Approximate Bayesian Computation in Large Scale Structure

ChangHoon Hahn†[1], Mohammadjavad Vakili†[1], Kilian Walsh[1], Andrew P. Hearin[2], David W. Hogg[1,3,4], and Duncan Campbell[5]

[1]Center for Cosmology and Particle Physics, *New York University*

Introduction

Cosmology was revolutionized in the 1990s with the introduction of likelihoods — probabilities for the data given the theoretical model— for combining data from different surveys and performing principled inferences of the cosmological parameters. Likelihood functions are easier to write down and evaluate when things are closer to Gaussian. However, in large-scale structure with galaxies, quasars, and quasar absorption systems as tracers the likelihood *cannot* be Gaussian.

The standard approach makes the strong assumption that the likelihood function for the data can be approximated by a Gaussian pseudo-likelihood function. Fortunately, this assumption is not required. There are principled, efficient methods such as Approximate Bayesian Computation, for minimizing computation and delivering correct posterior inferences

ABC-PMC

ABC provides a *rejection sampling* framework for inferring the posterior probability distribution using only a data simulator and some choices of summary statistics. It approximates $p(\vec{\theta}, X|D)$ by drawing proposals $\vec{\theta}$ from the prior over the model parameters, simulating the data from the proposals, and then rejecting the proposals that are beyond a certain threshold "distance" from the data. The distance metric, in principle, can be any positive definite function that compares the chosen summary statistics between the data and the simulation. In practice, we use ABC in conjunction with a more efficient sampling operation like Population Monte Carlo (PMC).

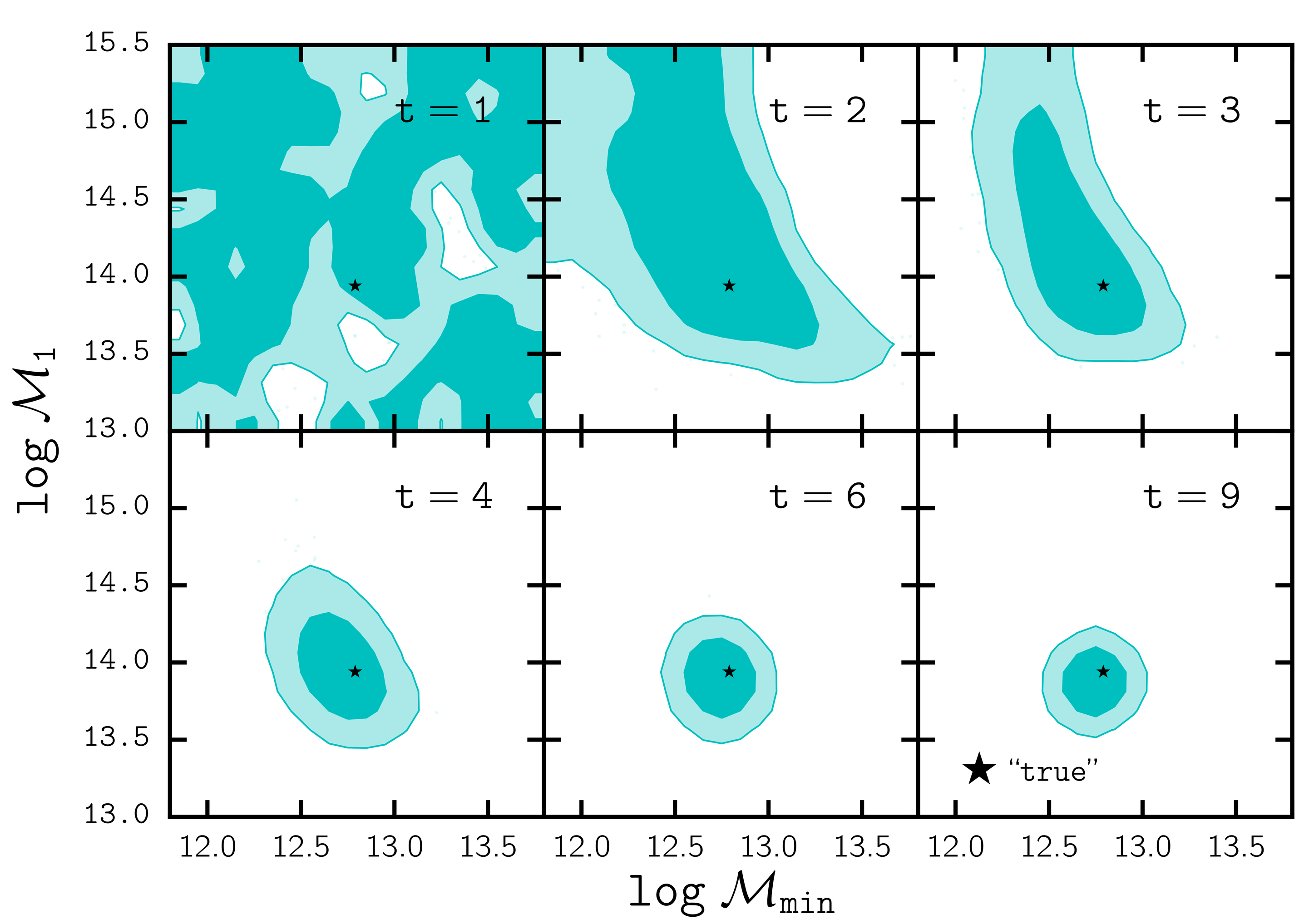

Figure: Evolution of the ABC particles in HOD parameter space. The black stars represent the "true" parameters of our mock observation. The parameter space of $\vec{\theta}_t$ initially shrinks dramatically until it converges to the posterior distribution of the parameters. At the final iteration, ABC has converged and $\vec{\theta}_{\rm true}$ lies safely within the 68% confidence region.

Halo Occupation Distribution

The assumption that galaxies reside in dark matter halos is the underlying bedrock of all contemporary theoretical predictions for galaxy clustering. The HOD characterizes this galaxy-halo connection. The central quantity in HOD is $p(N_{\rm g}|M_{\rm h})$, the probability that a halo of mass $M_{\rm h}$ hosts $N_{\rm g}$ galaxies. In the HOD model we use from Zheng et al.(2007):

$$\langle N_{\rm cen}\rangle = \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{\log M - \log M_{\rm min}}{\sigma_{\log M}}\right)\right], \quad \langle N_{\rm sat}\rangle = \langle N_{\rm cen}\rangle \left(\frac{M-M_0}{M_1}\right)^{\alpha}.$$

We use `Halotools` for our HOD forward model and for calculating summary statistics (number density, two-point correlation function, and group multiplicity function) of the simulated galaxy catalogs.

Results

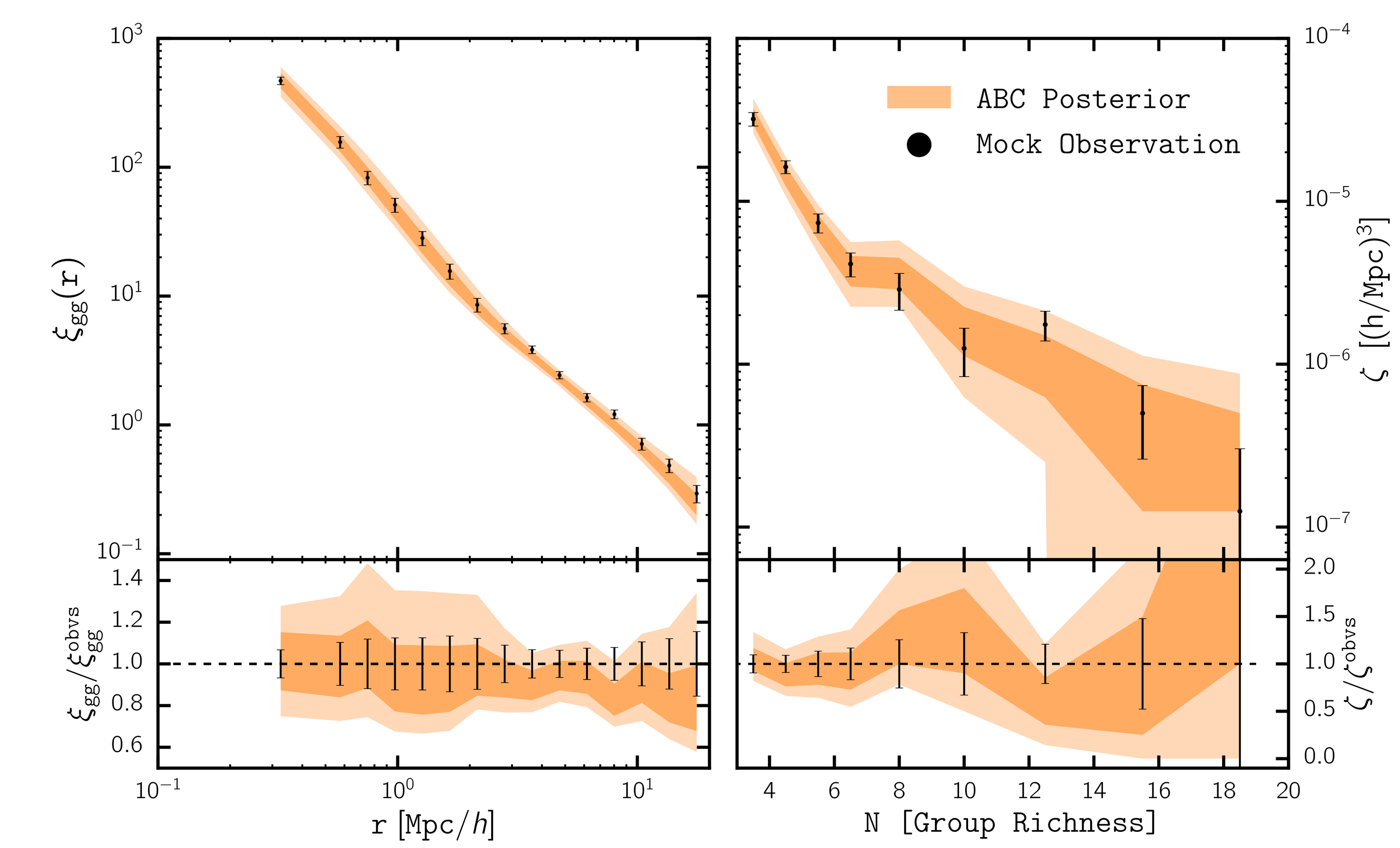

Figure: Comparison between the observables (ξ, left; $\zeta_{\rm g}$, right) predicted by the ABC-PMC posterior (orange) and the mock observation (black). The darker and lighter shaded regions represent the 68% and 95% confidence regions of the posterior predictions, respectively. The observables of the ABC-PMC posteriors and mock observations are in good agreement. The error-bars of the mock observations lie within the 68% confidence interval of the ABC-PMC posterior predictions.

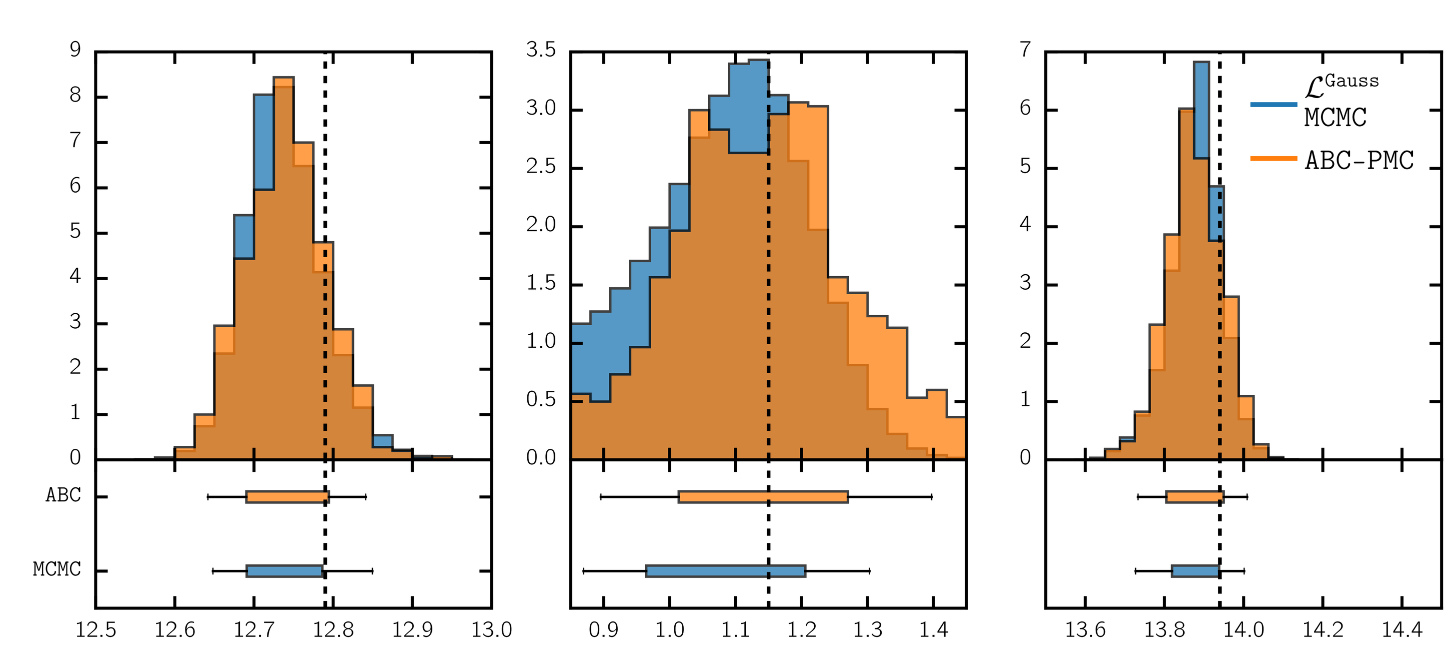

Figure: The marginalized posterior PDFs over the HOD parameter constraints from ABC-PMC (orange) and Gaussian Likelihood MCMC (blue). The "true" HOD parameters of our mock observations are represented with vertical black dashed line. Marginalized posterior PDFs obtained from the two methods are consistent with each other in both accuracy and precision.

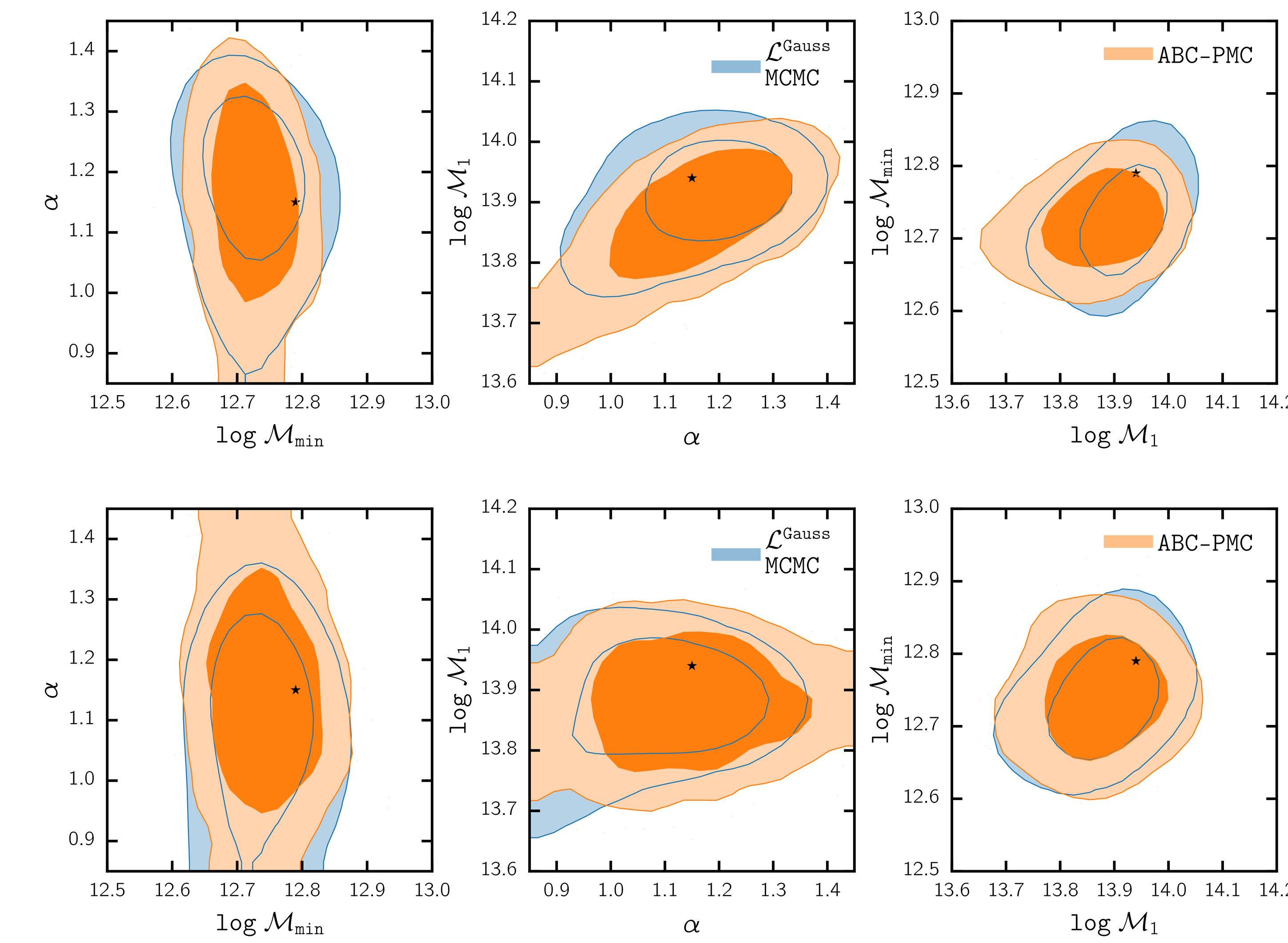

Figure: We compare the ABC-PMC (orange) and the Gaussian likelihood MCMC (blue) predictions of the 68% and 95% posterior confidence regions over HOD parameters using $\bar{n}_{\rm g}$ and $\zeta_{\rm g}(N)$ as observables. Black stars represent the "true" HOD parameters used to create the mock observations. Both approaches are consistent with the true values.

- The standard approaches to Bayesian parameter inference in LSS assume a Gaussian functional form for the likelihood. They are also typically restricted to measurements such as the two point correlation function.
- Likelihood free inferences such as Approximate Bayesian Computation make inference possible without assuming any functional form of the likelihood, thereby relaxing the standard assumptions and restrictions.
- We demonstrate that ABC, with Population Monte Carlo, is feasible for LSS parameter inference by using it to constrain parameters of the halo occupation distribution model for populating dark matter halos with galaxies in mock observations.
- The HOD parameter constraints from our ABC implementation are consistent with the constraints from the standard approach (pseudo-likelihood function of Gaussian form with MCMC).

Ultimately our results suggest that *ABC can be applied in parameter inference for LSS analyses.*

`chh327@nyu.edu`

†These authors have contributed equally to the work.

[2] Yale Center for Astronomy & Astrophysics

[3] Max-Planck-Institut für Astronomie, Heidelberg, Germany

[4] Department of Astronomy, Yale University